

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/333511648>

AN EVALUATION OF LANDSLIDE SUSCEPTIBILITY MAPPING USING REMOTE SENSING DATA AND MACHINE LEARNING ALGORITHMS IN IRAN

Conference Paper · May 2019

DOI: 10.5194/isprs-annals-IV-2-W5-503-2019

CITATION

1

READS

198

6 authors, including:



Bahareh Kalantar

61 PUBLICATIONS 545 CITATIONS

[SEE PROFILE](#)



Mohamed Barakat A. Gibril

University of Sharjah

15 PUBLICATIONS 68 CITATIONS

[SEE PROFILE](#)



Usman Lay

Universiti Putra Malaysia

10 PUBLICATIONS 25 CITATIONS

[SEE PROFILE](#)



Alireza Motevalli

Tarbiat Modares University

16 PUBLICATIONS 230 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Multiple Moving Object Detection From UAV Videos Using Trajectories of Matched Regional Adjacency Graphs [View project](#)



earthquake- induced building damage detection [View project](#)

AN EVALUATION OF LANDSLIDE SUSCEPTIBILITY MAPPING USING REMOTE SENSING DATA AND MACHINE LEARNING ALGORITHMS IN IRAN

B. Kalantar ^{1,*}, N. Ueda ¹, H. A. H. Al-Najjar ², M. B. A. Gibril ³, U. S. Lay ⁴, A. Motevalli ⁵

¹RIKEN Center for Advanced Intelligence Project, Goal-Oriented Technology Research Group, Disaster Resilience Science Team, Tokyo 103-0027, Japan – ([bahareh.kalantar](mailto:bahareh.kalantar@riken.jp), [naonori.ueda](mailto:naonori.ueda@riken.jp))@riken.jp

²Centre for Advanced Modelling and Geospatial Information Systems (CAMGIS), Faculty of Engineering and IT, University of Technology Sydney, 2007 NSW, Australia – HusamAbdulrasoolHammadeh.AL.NAJJAR@student.uts.edu.au

³Research Institute of Sciences and Engineering, University of Sharjah, Sharjah 27272, UAE - mbgibril@sharjah.ac.ae

⁴Department of Civil Engineering, Faculty of Engineering, Universiti Putra Malaysia, Serdang, Selangor, Malaysia - usmansirly@gmail.com

⁵Department of Watershed Management Engineering, College of Natural Resources, Tarbiat Modares University, Noor, Mazandaran, Iran - motevalli.alireza@yahoo.com

KEY WORDS: Machine Learning, Landslide, Random Forest, Decision Tree, Boosted Regression Tree

ABSTRACT:

Landslide is painstaking as one of the most prevalent and devastating forms of mass movement that affects man and his environment. The specific objective of this research paper is to investigate the application and performances of some selected machine learning algorithms (MLA) in landslide susceptibility mapping, in Dodangeh watershed, Iran. A 112 sample point of the past landslide, occurrence or inventory data was generated from the existing and field observations. In addition, fourteen landslide-conditioning parameters were derived from DEM and other topographic databases for the modelling process. These conditioning parameters include total curvature, profile curvature, plan curvature, slope, aspect, altitude, topographic wetness index (TWI), topographic roughness index (TRI), stream transport index (STI), stream power index (SPI), lithology, land use, distance to stream, distance to the fault. Meanwhile, factor analysis was employed to optimize the landslide conditioning parameters and the inventory data, by assessing the multi-collinearity effects and outlier detections respectively. The inventory data is divided into 70% (78) training dataset and 30% (34) test dataset for model validation. The receiver operating characteristics (ROC) curve or area under curve (AUC) value was used for assessing the model's performance. The findings reveal that TRI has 0.89 collinearity effect based on variance-inflated factor (VIF) and based on Gini factor optimization total curvature is not significant in the model development, therefore the two parameters are excluded from the modelling. All the selected MLAs (RF, BRT, and DT) shown promising performances on landslide susceptibility mapping in Dodangeh watershed, Iran. The ROC curve for training and validation for RF are 86% success rate and 83% prediction rate implies the best model performance compared to BRT and DT, with ROC curve of 72% and 70% prediction rate, respectively. In conclusion, RF could be the best algorithm for producing landslide susceptibility map, and such results could be adopted for the decision-making process to support land use planner for improving landslide risk assessment in similar environmental settings.

1. INTRODUCTION

Globally, man encounter challenges in resolving check and balances between the search for shelter and the growing risk of environmental hazard as a result of climate change. The most densely populated regions around the world are the hills and mountainous areas, which are often prone to numerous forms of natural hazard including landslides. Landslide is a complex gravitational flow that initiates mass movements. Different categories of landslides and associated triggering factors have been reported in the literature with ambiguous definitions (Lollino et al. 2015; Pradhan et al. 2017a).

A number of models with diverging steps of generalization have been established and investigated in geospatial science for evaluating landslide susceptibility. The models are categorized into five classes (Pradhan et al. 2017b), which include bivariate statistic, multivariate statistic, expert-based, machine learning and hybrid models. The individual models are subdivided into numerous subcategories with holds associated merits and demerits.

Applications of these model groups in landslide mapping are reported in the literature these include analytic hierarchy process (AHP) (Sharma and Mahajan 2018) that requires expert knowledge, and fuzzy analytic hierarchy process (FAHP) (Yang et al. 2017) have incorporated expert-based model, which inventory data is not a requisite in the learning process. Nevertheless, the decision on the contributions of the landslides conditioning variables is subjective.

Other landslides analytical models subclass like the weight of evidence (WOE) (Ilija and Tsangaratos 2016), statistical index (SI) (Razavizadeh et al. 2017), and frequency ratio (FR) (Sharma and Mahajan 2018) fits in the bivariate statistical models. While the logistic regression (LR) (Du et al. 2018), discriminate analysis (DA) (Pham and Prakash 2017), partial least squares (PLS) regression (Pradhan et al. 2017b) are multivariate statistical models group. The bivariate numerical models determine the impact of landslides conditioning variables on the menace incidence although, the focal drawback of this approach is the notion of conditional objectivity. On the other hand, the multivariate numerical models investigate the

* Corresponding author

relationship and involvements of all conditioning variables in predicting the occurrence of landslides; the perceived weakness associated with the models include huge data demand over a large scales regarding landslide distribution.

Different MLA have rapidly advanced in recent time these include support vector machine (SVM) (Kalantar et al. 2018a; Piragnolo et al. 2017; Feizizadeh et al. 2017; Lin et al. 2016); artificial neural network (ANN) (Kalantar et al. 2018b; Dou et al. 2017), decision tree (DT) (Pradhan 2013; Wang et al. 2016), boosted regression tree (BRT) (Lombardo et al. 2015), and random forest (RF) (Chen et al. 2018; Zhang et al. 2017; Pirotti et al. 2016; Paudel et al. 2016). Studies have shown that MLA models interpret the nonlinear relationship, handles missing values with better analytical capability than conventional regression approach, and apply uncertainty in landslide inventory. The most promising MLA indicate effective for landslide susceptibility mapping area SVM, RF, and ANN. However, they are time costing and have difficulty in detecting uncertainty directly, in view of these there is a need to improve the approaches considerably.

Lastly, to astound various potential drawbacks emanate from individual models, hybrid models is developed; these comprised FR-SVM, FR-LR, and WoE-RF, AHP-FR (Yan et al. 2019), a combined fuzzy and support vector machine (F-SVM) (Meng et al. 2016), integration of support vector machine and random space (Hong et al. 2017). Often report high complexity in the integrated models. Selection of most appropriate approach is one of the challenges for land use planner.

Therefore, to develop a landslide susceptibility models three MLA were explored and evaluated, they have been considered effective prediction tools in dealing with dependent and explanatory variables of non-linear relationships. These include a random forest (RF), decision tree (DT), and boosted regression tree (BRT) methods. The specific objective of this study is to investigate the optimum prediction model from the advanced MLA that is suitable for landslide susceptibility in Dodangeh watershed, the province of Mazandaran, Iran. For this purpose, 112 landslide inventories point data and 14 landslide conditioning variables derived from digital elevation model (DEM) with geographic information system (GIS) tools and other topographic databases were prepared.

Lastly, the viability and success of the adopted MLA have been evaluated and compared using performance metrics. The landslide inventory dataset is divided into the training (70%) and the test (30%) data. The best model attainment from the proposed models in this study will contribute the urban engineers in the decision-making process in the area of land use allocations and suitability for risk-free zones.

2. STUDY AREA

The Dodangeh District is located in the Sari County, province of Mazandaran, Iran with a geographic position on latitude $36^{\circ}24'44.56''\text{N}$ and longitude $53^{\circ}14'34.78''\text{E}$ was considered as the study area. The area suffered seriously of landslides effects during the wet season and have a maximum elevation of 2800m with a population size of 8,140 in 2006. Two climatic seasons (dry and rainy) exist in the area. The vegetation cover consists of forest, agriculture, orchard and mixture of both (Figure 1a). The lithology structure of the area is shown in Figure 1b.

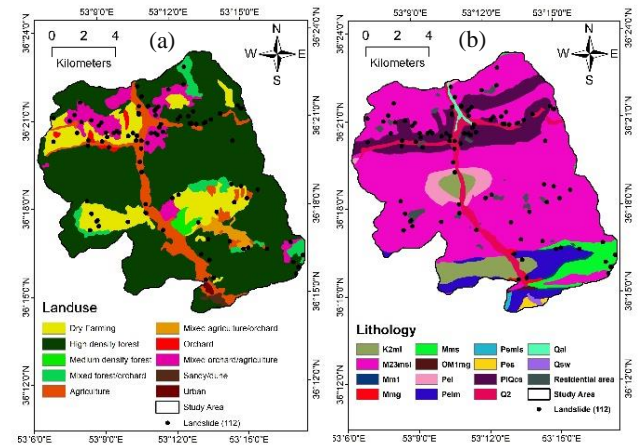


Figure 1. (a) Landuse and (b) Lithology maps of the study area.

3. MATERIALS AND METHODS

The methodological flow chart in Figure 2 illustrated five steps cover this research paper; including (i) landslide conditioning factors, (ii) landslide inventories, (iii) factor analysis and optimization (iv) machine learning methods (RF, DT, BRT), (v) validation using area under curve (AUC).

3.1 Landslide Conditioning Factors

In this study, fourteen conditioning factors, which includes TWI, TRI, total curvature, STI, SPI, slope, profile curvature, plan curvature, lithology, land use, distance to stream, distance to fault, altitude, aspect were derived from DEM and other topographic databases. While altitude, aspect, slope, curvature (profile, plan, total) were derived from DEM of 10m resolution, and TWI, TRI, STI, SPI were calculated using the flowing formula:

$$\text{SPI} = A_s \tan \beta \quad (1)$$

$$\text{TWI} = \log_e \left(\frac{A_s}{\tan \beta} \right), \quad (2)$$

$$\text{STI} = \left(\frac{A_S}{22.1} \right)^{0.6} \left(\frac{\sin \beta}{0.0896} \right)^{1.3}, \quad (3)$$

$$\text{TRI} = \sqrt{[x](\max^2 + \min^2)} \quad (4)$$

Where A_s = area of catchment (m^2)
 β = gradient of the slope in radians (Hong et al. 2018)
 max, min = largest and minimum value of pixel in
 nine rectangular altitude neighbourhoods
 (Kalantar et al. 2018a).

Distance to faults and stream are generated using Euclidean distance function in ArcMap. Furthermore, the variables were classified using quantile range in ArcGIS software environment, altitude was reclassified into five classes (210-420m, 430-560m, 570-700m, 710-840m, 850-1200m), the slope angle was reclassified into five classes (0-7.3°, 7.4-12°, 13-17°, 18-24°, 25-67°) (Figure 3a and 3b), slope aspect was classified into nine classes of directions (flat, Northeast, East, Southeast, South, Southwest West, Northwest, and North Figure 3c). In addition, the profile and plan curvature were categorized in three classes convex (negative values), flat (zero

value) and concave (positive values) (Figure 3d and 3e). The land cover classes are dry farming, highly dense forest, mixed forest and orchard, agriculture, mixed agriculture and orchard, orchard, mixed orchard and agriculture, Sandy/dune and urban area were illustrated in (Figure 2a). Lithology was used as mentioned in previous section (Figure 2b). However, the TWI, STI, and SPI were ordered into five classes as shown in Figure 3f, 3g, and 3h, respectively. Finally, Distance to stream and Distance to fault were categorized into five classes, see (Figure 3i and 3j).

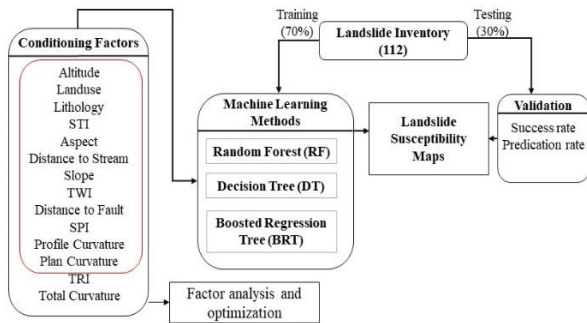


Figure 2. Overall workflow of this study.

3.2 Landslide Inventories

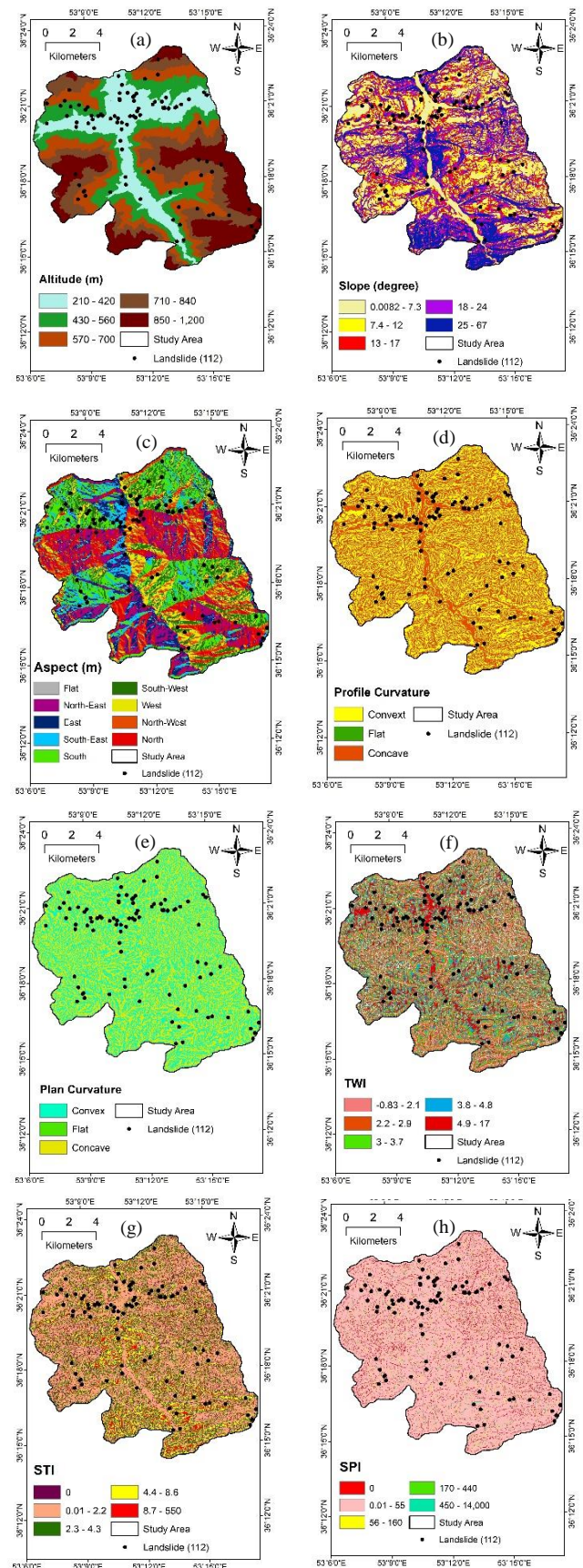
112 randomly landslide inventories points have been obtained with 14 different landslides conditioning variables (independent variables). The inventory is sourced from visual interpretation, previous reports, aerial photographs and satellite images. 70% of the landslide inventories were used for training and 30% was used for testing (Hong et al. 2018).

In a successful prediction model, there must be a dependent variable (y) and independent or predictors variable (x_1, \dots, x_n). To develop the landslides susceptibility maps, we considered two rejoinders (landslide denoted by 1 and no landslide denoted by 0) as the (y) variable and the conditional variable as x -variables. The generated datasets are continuous in natures and subjected to pre-processing that involve a check for multicollinearity effects, out layer evaluation feature contribution. Although, the variables were normalized prior to the model development to get rid of large dissimilarity and concentration on the certain variable principal dispersals. In view of this, variable scale domination of different variables; then the dataset has a unit (1) variance and zero (0) mean through individual conditioning variable. Hence, z-score normalization was utilized, that could present as follows:

$$\text{norm}, X_{ij} = \frac{X_{ij} - \text{mean } X_j}{\text{std } X_j} \quad (5)$$

3.3 Statistical Analysis and Optimization

The next step is factor analysis was applied in the pre-process section to assess for the presence and eliminate of collinearity effect in a given landslide conditioning parameters and outlier values in the inventory dataset (Pradhan et al. 2017b). Multi-collinearity denotes signifies the existence of a strong relationship amongst the independence or conditioning variable with one another in the model. Note that the presence of multi-collinearity reduces the model performance by increasing the error term.



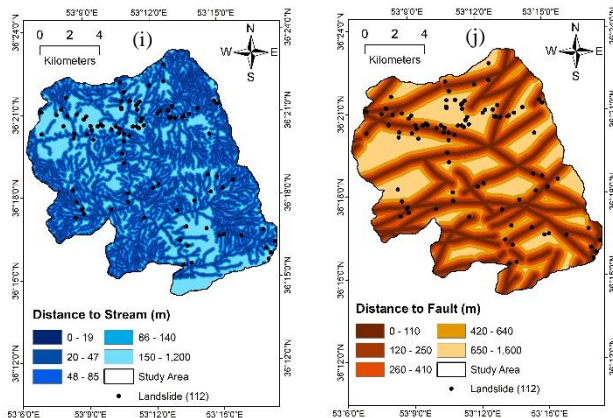


Figure 3. Conditioning factors.

Researchers have developed and recommended a various approach to address multi-collinearity such as (a) discarding highly correlated features (b) Linear combination of the highly correlated features and (c) implementing progressive simulations, which explain multi-collinearity effect (Pradhan et al. 2017b). We adopted the highly related features discard approach using an estimation of variance-inflated factor (VIF) as the following equation:

$$VIF = \frac{1}{1 - R^2} \quad (6)$$

where R^2 = multi correlation coefficient between individual feature and the other features in the model.

In the current study, the factors with a VIF greater than 4.00 is removed. Table 1 displays the estimated VIF values. The VIF values show that the TRI (VIF = 4.16) is the factor which suffers multi-collinearity in the dataset. Therefore, TRI is discarded (Pradhan et al. 2017b).

Factor optimization is another important stage in landslide susceptibility mapping. A large number of conditioning factors could increase the training sample size and computational cost. Consequently, the estimated regression coefficients are misled when the number of factors increased. In this research, attributes of factor optimization that include *Chi-square* and Gini importance was applied to identify the important features and get rid of the insignificance variables at 0.05 (95%) confidence level, used for further analysis. Table 2, shows the results of the model input selected based on their significant importance. Accordingly, the factors optimizations' attributes have all agreed in selecting three factors (altitude, land use and lithology), which are considered the most important features for the prediction of landslide susceptibility in Dodangeh watershed.

Factor	Multi-collinearity and Statistics summary			
	Means	Standard Dev.	Multiple	VIF
TWI	2.71	1.63	0.66	1.78
TRI	2.77	1.39	0.87	4.16
Total Curvature	1.94	1.00	0.62	1.63
STI	2.61	1.07	0.73	2.19
SPI	1.96	0.57	0.49	1.33
Slope	3.02	1.39	0.85	3.71
Profile Curvature	2.09	0.99	0.37	1.16
Plan Curvature	2.02	1.00	0.49	1.32

Lithology	8.33	3.01	0.26	1.07
Land use	2.13	1.60	0.18	1.03
Distance to Stream	103.67	128.25	0.11	1.01
Distance to Fault	416.40	311.96	0.07	1.00
Altitude	2.62	1.4801	0.62	1.62
Aspect	185.26	103.59	0.14	1.02

Table 1. The estimated Variance Information Factor (VIF) for landslide conditioning variables.

Variables	Chi-square (X^2) method		Gini Method		
	(X^2)	p-values	Gini	IV	Cramer's V
Altitude	69.284	0.000	0.354	1.424	0.539
Land use	63.483	0.000	0.367	1.188	0.514
Lithology	35.529	0.000	0.463	0.319	0.271
TRI	21.823	0.000	0.462	0.313	0.274
STI	14.162	0.006	0.465	0.065	0.262
Aspect	11.802	0.160	0.465	0.284	0.261
Distance to Stream	10.892	0.091	0.477	0.188	0.211
Slope	9.336	0.053	0.479	0.180	0.204
TWI	4.852	0.434	0.485	0.119	0.170
Distance to Fault	4.754	0.689	0.489	0.032	0.146
SPI	2.813	0.244	0.493	0.050	0.111
Profile	1.460	0.226	0.496	0.026	0.080
Curvature Plan	0.017	0.893	0.499	0.0003	0.008
Total Curvature	0.000	1.000	0.500	0.000	0.000

Table 2. The factors importance based on factor optimization (Chi-square and Gini).

3.4 Machine Learning Algorithm (MLA)

This study employed three MLAs (RF, DT and BRT) in developing the landslide susceptibility models using R 3.0.2 (an open source software). RF approach is an ensemble machine learning method that creates a large amount of DT; used to describe the spatial relationship that exists in landslide events. In contrast to other algorithms, RF has different procedures for essential factors. A recommended procedure is the influence on the classification accuracy because the value of the factor in a developed tree-like structure was evaluated randomly (Pradhan et al. 2017b). Application of RF method in landslides susceptibility has used the gain of high modification among the different trees, which permit individual tree to choose a class association or membership and allocation groups of a particular class are based on highest number of the supported polls. This ensemble exhibited a promising performance on complex datasets, with a simple preparation and processes (Stumpf and Kerle 2011).

The second approach is the DTs, which have reported to be an effective data-mining tool capable of predicting both continuous and categorical (discrete) response features (Lee et al. 2006). Integrated nodes in the model are splits into different observation. Existing components of the initial nodes begin to nurture a tree via the training data and the explanatory features resulted in a number of division of child nodes. It is possible for the algorithm to further produce more division from the child nodes but advance split is not allowed at the node terminal. The model prediction ability depends on the structure of the tree nodes terminal; this procedure has numerous advantages. The most paramount importance of the model is the ability to predict

complex relationships between variables and give a simple explanation of a DT (Bui et al. 2012). Moreover, complex mathematical equations are not associated with DTs model for any predictions. Finally, the demerits attributed to this model involves their proneness to filthy data and allowed limited attributes result. for more details on DT equations, the interested reader is directed to (Bui et al. 2012).

BRT is a hybrid of statistical and machine learning methods, aimed at improving the performance of a model through appropriate integration of various models for prediction of an event (Schapire 2003). The latest development in BRT approach is demonstrated in modelling the occurrences of natural hazard that is not linear in nature and prove adequate in conducting complex nonlinear relationships. In-depth feature or data, pre-processing (transformation and outlier detection) is not necessary for this approach because the correlation effects between variables are reported automatically (Elith et al. 2008). The two approaches that made up BRT are regression and boosting, which produces intuitive results represented in a simple visualize on DTs. Some outstanding characteristics of the DT models exist in literature, which includes, unaffected to outlier and surrogate approach is used to amend missing data in model input variables (Elith et al. 2008; Breiman et al. 1984). The boosting algorithm purposely employed to enhances model performance accuracy because it is capable of detecting numerous irregular procedural search better than to discover a solitary optimal or extreme forecasting control instruction (Schapire 2003). Addition of several appropriate trees in the BRT model will surmount the model's weakness.

3.5 Validation of the Model's Performance

The initial analyses consist of the ability to create successfully landslide susceptibility models emphasis on address the highly susceptible area, this so-called success rate. The prediction rates are the ability of the test dataset to assess accurately the performances of the models' prediction strength. Note that, a landslide is not evenly distributed in the area. There is not concrete thumb rule regarding the ratio of training and test dataset. Hence, in this research, the landslides inventory was divided into training (70%) and test (30%) set. The popular approach to decide the accurate performance of analytical investigation (Razavi Termeh et al. 2018). In this study models performance was conducted using receiver operating characteristic (ROC) curves (AUC) were considered (Pourghasemi et al. 2013); using success rate and prediction rate for the assessments of the robustness of individual machine learning algorithms in landslide susceptibility mapping (Pham et al. 2018). The ROC (AUC) value is categorized into scales in relation to qualitative classes 50% to 60% is poor, 60% to 70% is average, 70% to 80% is good, 80% to 90% is very good and 90% to 100% is exceptional (excellent) (Razavi Termeh et al. 2018; Yesilnacar and Topal 2005).

4. RESULT AND DISCUSSION

4.1 Relative Importance Analysis of Landslide Conditioning Parameters

Concerning the predictors' contribution in the model development, *chi-square* results in Figure 4, shows that altitude (DEM) (69.28%), Landuse (63.48%) were the most contributing parameters, next by Lithology (36.53%), TRI (22.82%), STI (14.16%), Aspect (12.80%), Distance to stream (10.89%), the minimum importance conditioning parameters were slope (9.34%), TWI (4.85%), Distance to the fault (4.75%), SPI (2.81%), Profile Curvature (1.46%), Plan curvature (0.02%) and no significance conditioning parameters is Total curvature (0.00%). Therefore, all these conditioning parameters expect total curvature were utilized as thematic

layers dataset were selected as input predictors in generating the landslide susceptibility maps. Since the analytical report of *chi-square* revealed their participation in the landslide prediction in Iran.

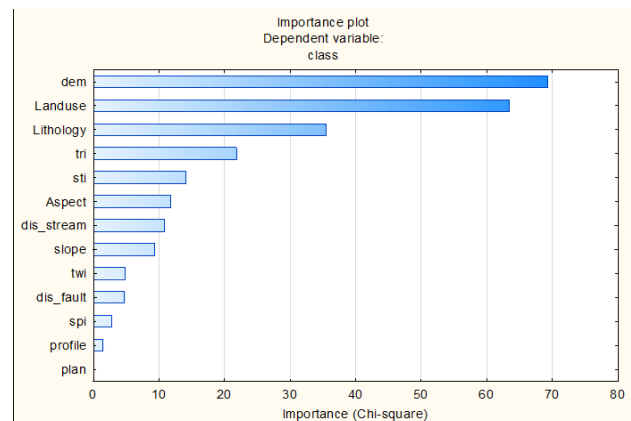


Figure 4. The important plot of conditioning factors using Chi-square.

The emergence of altitude and land use as the most contributing conditioning parameters to landslide occurrence; justifies the logical anthropogenic activities, which has backward effects on the natural environments and manifested on the altitude. These factors were proved leading factor capable to accelerate landslide activities in the study area.

Factor analysis showed that the TRI with VIF values of 4.16 is the highly correlated factors and need to be removed. On the other hand, results of factor optimization indicated that total curvature with *chi-square* and information value equal to zero was found to be statistically not significant. Important factors in predicting landslides in Dodangeh watershed which are altitude (Chi-square value = 69.284, information value = 1.424), land use (Chi-square value = 63.483, information value = 1.188), and lithology (Chi-square value = 35.529, information value = 0.319). On the other hand, total curvature (Chi-square value = 0.00, information value = 0.00) is removed as its important value is zero.

4.2 The Coefficient of the Conditioning Parameter

The developed coefficient for each factor is shown in Table 3, from different machine learning algorithms. The RF algorithm estimates values range from 0 to 1; altitude reveal to yield the highest estimated value of 1 and the lowest coefficient is 0.073 for plan curvature. Meanwhile, the DT generated coefficient results that presented land use factor with the highest coefficient value of 1, the conditioning factor differs from the RF. It is observed that at the bottom estimates for both DT and RF models, shown a consistency in their coefficient results, though DT minimum coefficient values are 0.044 on plan curvature. Finally, the BRT highest coefficient result is similar to the RF model on the same conditioning parameters altitude. Hence, the developed output model using the RF, DT and BRT models are presented in Figure 5.

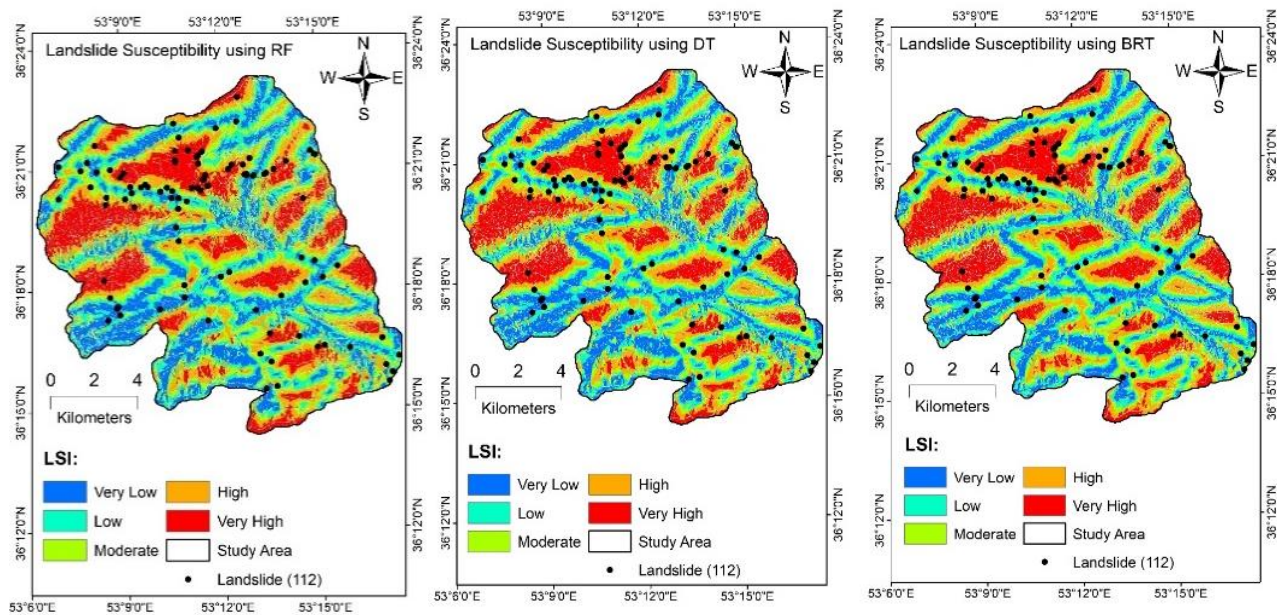


Figure 5: Landslide susceptibility maps produced by the RF, DT, and BRT.

Factor	Coefficients		
	RF	DT	BRT
Altitude	1.000	0.635	1.000
Landuse	0.811	1.000	0.614
Lithology	0.348	0.284	0.841
STI	0.370	0.258	0.374
Aspect	0.496	0.331	0.762
Distance to Stream	0.428	0.250	0.616
Slope	0.215	0.183	0.445
TWI	0.194	0.151	0.548
Distance to Fault	0.402	0.284	0.844
SPI	0.130	0.063	0.416
Profile Curvature	0.155	0.068	0.136
Plan Curvature	0.073	0.044	0.212

Table 3. Estimated coefficients of landslide conditioning factors by the RF, DT, and BRT.

4.3 Validation of the Models

The predictive and consistency strength of these approaches for landslide susceptibility was evaluated by success curve using the training data and prediction rate curves drawn from testing data that were not included in the training process) (Bui et al. 2018). The ROC approach used in this study owing to its reputation, satisfactory and efficiency to quantitatively estimates of models. The observed findings for this study on the success rate and prediction rate curve for the three machine learning approaches were shown in Figure 6. In comparison, the RF models proved to outperformed other machine learning

models with the success rate AUC value of 0.86 (86%) and prediction rate AUC value 0.83 (83%). While the DT model have corresponding AUC values for success rates and prediction rates of 0.72 and 0.70 respectively, and BRT have the AUC values for success rates and prediction rates of model 0.75, and 0.72 (Table 4).

Methods	Success rate	Prediction rate
RF	0.86 (86%)	0.83 (83%)
BRT	0.75 (75%)	0.72 (72%)
DT	0.72 (72%)	0.70 (70%)

Table 4. Success and predication rates for RF, DT, and BRT.

The prediction rate was derived from the (30%) testing dataset, which was used in the evaluation of the capability of the models in predicting landslides zones. However, the highest predictive model strength for landslide susceptibility areas in the Dodangeh watershed was evidence by the RF model follow by BRT and then DT (Figure 6). RF model emerged to be the global best model in both training rate and prediction rate for landslide modelling in the study area.

So far, the findings of this study have been compared with similar research in the same line. The results of this study agreed with the findings of Pradhan et al. (2017b) that investigated and compared SVM, RF and DT models. Their result revealed RF model the best compared to SVM and DT. There is a need for advanced research in environmental hazards now due to the rising human population resulting in high demand in shelter and infrastructures; that require an accurate report on the hazard risk-free or suitable area setting.

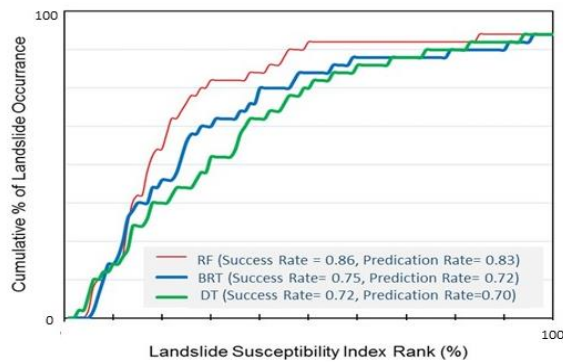


Figure 6: Success and prediction rates for the machine learning models.

5. CONCLUSION

This research evaluated and verified machine learning (RF, DT, and BRT) models in landslide susceptibility mapping with a case study at Dodangeh watershed, Mazandaran province, Iran. A GIS database model was created using 112 sample points of landslide inventory coupled with 12 landslides conditioning parameters. The AUC values for the success rates and prediction rate was detailed from the model's validations.

Research findings indicated RF model to have the best AUC values of success rate (0.86) and prediction rate (0.83) rate, following BRT model (success=0.75, prediction=0.72) and DT model (success=0.72, prediction=0.70). The results displayed efficacy of the three machine learning models to predict landslide susceptibility maps with significant accuracy. The research is relevant to reliable urban planning managers and engineers, as a tool for the adequate decision-making process.

References

- Breiman, L., J. H. Friedman, R. A. Olshen, and C. J. Stone. 1984. "Classification and Regression Trees." Monterey, Calif., USA: Wadsworth, Inc.
- Bui, Dieu Tien, Mahdi Panahi, Himan Shahabi, Vijay P. Singh, Ataollah Shirzadi, Kamran Chapi, Khabat Khosravi, et al. 2018. "Novel Hybrid Evolutionary Algorithms for Spatial Prediction of Floods." *Scientific Reports* 8 (1). Nature Publishing Group: 15364. doi:10.1038/s41598-018-33755-7.
- Bui, Dieu Tien, Biswajeet Pradhan, Owe Lofman, Inge Revhaug, and Oystein B. Dick. 2012. "Application of Support Vector Machines in Landslide Susceptibility Assessment for the Hoa Binh Province (Vietnam) with Kernel Functions Analysis." *International Environmental Modelling and Software Society (iEMSs)*. doi:10.13140/RG.2.1.3041.4889.
- Chen, Wei, Jianbing Peng, Haoyuan Hong, Himan Shahabi, Biswajeet Pradhan, Junzhi Liu, A. Xing Zhu, Xiangjun Pei, and Zhao Duan. 2018. "Landslide Susceptibility Modelling Using GIS-Based Machine Learning Techniques for Chongren County, Jiangxi Province, China." *Science of the Total Environment* 626 (June). Elsevier: 1121–1135. doi:10.1016/j.scitotenv.2018.01.124.
- Dou, Jie, Hiromitsu Yamagishi, Zhongfan Zhu, Ali P. Yunus, and Chi Wen Chen. 2017. "TXT-Tool 1.081-6.1 A Comparative Study of the Binary Logistic Regression (BLR) and Artificial Neural Network (ANN) Models for GIS-Based Spatial Predicting Landslides at a Regional Scale." In *Landslide Dynamics: ISDR-ICL Landslide Interactive Teaching Tools: Volume 1: Fundamentals, Mapping and Monitoring*, 139–151. Cham: Springer International Publishing. doi:10.1007/978-3-319-57774-6_10.
- Du, Guoliang, Yongshuang Zhang, Zhihua Yang, Changbao Guo, Xin Yao, and Dongyan Sun. 2018. "Landslide Susceptibility Mapping in the Region of Eastern Himalayan Syntaxis, Tibetan Plateau, China: A Comparison between Analytical Hierarchy Process Information Value and Logistic Regression-Information Value Methods." *Bulletin of Engineering Geology and the Environment*, October 4. doi:10.1007/s10064-018-1393-4.
- Elith, J., J. R. Leathwick, and T. Hastie. 2008. "A Working Guide to Boosted Regression Trees." *Journal of Animal Ecology*. John Wiley & Sons, Ltd (10.1111). doi:10.1111/j.1365-2656.2008.01390.x.
- Feizizadeh, Bakhtiar, Majid Shadman Roodposhti, Thomas Blaschke, and Jagannath Aryal. 2017. "Comparing GIS-Based Support Vector Machine Kernel Functions for Landslide Susceptibility Mapping." *Arabian Journal of Geosciences* 10 (5). Springer Berlin Heidelberg: 122. doi:10.1007/s12517-017-2918-z.
- Hong, Haoyuan, Junzhi Liu, A. Xing Zhu, Himan Shahabi, Binh Thai Pham, Wei Chen, Biswajeet Pradhan, and Dieu Tien Bui. 2017. "A Novel Hybrid Integration Model Using Support Vector Machines and Random Subspace for Weather-Triggered Landslide Susceptibility Assessment in the Wuning Area (China)." *Environmental Earth Sciences* 76 (19). Springer Berlin Heidelberg: 652. doi:10.1007/s12665-017-6981-2.
- Hong, Haoyuan, Biswajeet Pradhan, Maher Ibrahim Sameen, Bahareh Kalantar, Axiang Zhu, and Wei Chen. 2018. "Improving the Accuracy of Landslide Susceptibility Model Using a Novel Region-Partitioning Approach." *Landslides* 15 (4). Springer Berlin Heidelberg: 753–772. doi:10.1007/s10346-017-0906-8.
- Ilija, Ioanna, and Paraskevas Tsangaratos. 2016. "Applying Weight of Evidence Method and Sensitivity Analysis to Produce a Landslide Susceptibility Map." *Landslides* 13 (2): 379–397. doi:10.1007/s10346-015-0576-3.
- Kalantar, Bahareh, Biswajeet Pradhan, Seyed Amir Naghibi, Alireza Motevalli, and Shattri Mansor. 2018a. "Assessment of the Effects of Training Data Selection on the Landslide Susceptibility Mapping: A Comparison between Support Vector Machine (SVM), Logistic Regression (LR) and Artificial Neural Networks (ANN)." *Geomatics, Natural Hazards and Risk* 9 (1). Taylor & Francis: 49–69. doi:10.1080/19475705.2017.1407368.
- Kalantar, Bahareh, Naonori Ueda, Husam Abdulrasool H. Mohammed Oludare Idrees, Alireza Motevalli, and Biswajeet Pradhan. 2018b. "Landslide Susceptibility Mapping at Dodangeh Watershed, Iran Using LR and ANN Models in GIS." In *SPIE: Earth Resources and Environmental Remote Sensing/ GIS Applications IX*, edited by Ulrich Michel and Karsten Schulz, 10790:107901D–1–10. SPIE. doi:10.1117/12.2501473.
- Lee, Tian Shyug, Chih Chou Chiu, Yu Chao Chou, and Chi Jie Lu. 2006. "Mining the Customer Credit Using Classification and Regression Tree and Multivariate Adaptive Regression Splines." *Computational Statistics and Data Analysis* 50 (4). North-Holland: 1113–1130. doi:10.1016/j.csda.2004.11.006.
- Lin, Yongliang, Kewen Xia, Xiaoqing Jiang, Jianchuan Bai, and Panpan Wu. 2016. "Landslide Susceptibility Mapping Based on Particle Swarm Optimization of Multiple Kernel

- Relevance Vector Machines: Case of a Low Hill Area in Sichuan Province, China.” *ISPRS International Journal of Geo-Information* 5 (10). Multidisciplinary Digital Publishing Institute: 191. doi:10.3390/ijgi5100191.
- Lollino, Giorgio, Daniele Giordan, Giovanni B. Crosta, Jordi Corominas, Rafiq Azzam, Janusz Wasowski, and Nicola Sciarra. 2015. *Engineering Geology for Society and Territory – Volume 2: Landslide Processes*. Edited by Giorgio Lollino, Daniele Giordan, Giovanni B. Crosta, Jordi Corominas, Rafiq Azzam, Janusz Wasowski, and Nicola Sciarra. *Engineering Geology for Society and Territory - Volume 2: Landslide Processes*. Cham: Springer International Publishing. doi:10.1007/978-3-319-09057-3.
- Lombardo, L., M. Cama, C. Conoscenti, M. Märker, and E. Rotigliano. 2015. “Binary Logistic Regression versus Stochastic Gradient Boosted Decision Trees in Assessing Landslide Susceptibility for Multiple-Occurring Landslide Events: Application to the 2009 Storm Event in Messina (Sicily, Southern Italy).” *Natural Hazards* 79 (3). Springer Netherlands: 1621–1648. doi:10.1007/s11069-015-1915-3.
- Meng, Qingkai, Fang Miao, Jing Zhen, Xinyuan Wang, An Wang, Ying Peng, and Qiang Fan. 2016. “GIS-Based Landslide Susceptibility Mapping with Logistic Regression, Analytical Hierarchy Process, and Combined Fuzzy and Support Vector Machine Methods: A Case Study from Wolong Giant Panda Natural Reserve, China.” *Bulletin of Engineering Geology and the Environment* 75 (3). Springer Berlin Heidelberg: 923–944. doi:10.1007/s10064-015-0786-x.
- Paudel, Uttam, Takashi Oguchi, and Yuichi Hayakawa. 2016. “Multi-Resolution Landslide Susceptibility Analysis Using a DEM and Random Forest.” *International Journal of Geosciences* 7 (5). Scientific Research Publishing: 726–743. doi:10.4236/ijg.2016.75056.
- Pham, Binh Thai, and Indra Prakash. 2017. “Evaluation and Comparison of LogitBoost Ensemble, Fisher’s Linear Discriminant Analysis, Logistic Regression and Support Vector Machines Methods for Landslide Susceptibility Mapping.” *Geocarto International*, November. Taylor & Francis, 1–18. doi:10.1080/10106049.2017.1404141.
- Pham, Binh Thai, Ataollah Shirzadi, Dieu Tien Bui, Indra Prakash, and M. B. Dholakia. 2018. “A Hybrid Machine Learning Ensemble Approach Based on a Radial Basis Function Neural Network and Rotation Forest for Landslide Susceptibility Modeling: A Case Study in the Himalayan Area, India.” *International Journal of Sediment Research* 33 (2). Elsevier: 157–170. doi:10.1016/j.ijsrc.2017.09.008.
- Piragnolo, Marco, Andrea Masiero, and Francesco Pirotti. 2017. “Open source R for applying machine learning to RPAS remote sensing images.” *Open Geospatial Data, Software and Standards* 2, no. 1 (2017): 16.
- Pirotti, F., Sunar, F. and Piragnolo, M., 2016. “Benchmark of machine learning methods for classification of a SENTINEL-2 images.” *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, 41.
- Pourghasemi, Hamid Reza, Biswajeet Pradhan, Candan Gokceoglu, Majid Mohammadi, and Hamid Reza Moradi. 2013. “Application of Weights-of-Evidence and Certainty Factor Models and Their Comparison in Landslide Susceptibility Mapping at Haraz Watershed, Iran.” *Arabian Journal of Geosciences* 6 (7). Springer-Verlag: 2351–2365. doi:10.1007/s12517-012-0532-7.
- Pradhan, Ananta Man Singh, Hyo Sub Kang, Saro Lee, and Yun Tae Kim. 2017a. “Spatial Model Integration for Shallow Landslide Susceptibility and Its Runout Using a GIS-Based Approach in Yongin, Korea.” *Geocarto International* 32 (4): 420–441. doi:10.1080/10106049.2016.1155658.
- Pradhan, Biswajeet. 2013. “A Comparative Study on the Predictive Ability of the Decision Tree, Support Vector Machine and Neuro-Fuzzy Models in Landslide Susceptibility Mapping Using GIS.” *Computers and Geosciences* 51 (February). Pergamon: 350–365. doi:10.1016/j.cageo.2012.08.023.
- Pradhan, Biswajeet, Maher Ibrahim Seeni, and Bahareh Kalantar. 2017b. “Performance Evaluation and Sensitivity Analysis of Expert-Based, Statistical, Machine Learning, and Hybrid Models for Producing Landslide Susceptibility Maps.” In *Laser Scanning Applications in Landslide Assessment*, 193–232. Cham: Springer International Publishing. doi:10.1007/978-3-319-55342-9_11.
- Razavi Termeh, Seyed Vahid, Aiding Kornejady, Hamid Reza Pourghasemi, and Saskia Keesstra. 2018. “Flood Susceptibility Mapping Using Novel Ensembles of Adaptive Neuro Fuzzy Inference System and Metaheuristic Algorithms.” *Science of the Total Environment* 615 (February). Elsevier: 438–451. doi:10.1016/j.scitotenv.2017.09.262.
- Razavizadeh, Samaneh, Karim Solaimani, Matteo Massironi, and Ataollah Kavian. 2017. “Mapping Landslide Susceptibility with Frequency Ratio, Statistical Index, and Weights of Evidence Models: A Case Study in Northern Iran.” *Environmental Earth Sciences* 76 (14). Springer Berlin Heidelberg: 499. doi:10.1007/s12665-017-6839-7.
- Schapire, Robert E. 2003. “The Boosting Approach to Machine Learning: An Overview.” In , 149–171. Springer, New York, NY. doi:10.1007/978-0-387-21579-2_9.
- Sharma, Swati, and A. K. Mahajan. 2018. “A Comparative Assessment of Information Value, Frequency Ratio and Analytical Hierarchy Process Models for Landslide Susceptibility Mapping of a Himalayan Watershed, India.” *Bulletin of Engineering Geology and the Environment*, March 9. doi:10.1007/s10064-018-1259-9.
- Stumpf, André, and Norman Kerle. 2011. “Object-Oriented Mapping of Landslides Using Random Forests.” *Remote Sensing of Environment*. doi:10.1016/j.rse.2011.05.013.
- Wang, Liang Jie, Min Guo, Kazuhide Sawada, Jie Lin, and Jinchu Zhang. 2016. “A Comparative Study of Landslide Susceptibility Maps Using Logistic Regression, Frequency Ratio, Decision Tree, Weights of Evidence and Artificial Neural Network.” *Geosciences Journal* 20 (1). The Geological Society of Korea: 117–136. doi:10.1007/s12303-015-0026-1.
- Yan, Fei, Qiuwen Zhang, Song Ye, and Bo Ren. 2019. “A Novel Hybrid Approach for Landslide Susceptibility Mapping Integrating Analytical Hierarchy Process and Normalized Frequency Ratio Methods with the Cloud Model.” *Geomorphology* 327 (February). Elsevier: 170–187. doi:10.1016/j.geomorph.2018.10.024.
- Yang, Weifeng, Zhihong Zheng, Xinquan Zhang, Baohua Tan, and Luyao Li. 2017. “Analysis of Landslide Risk Based on Fuzzy Extension Analytic Hierarchy Process.” *Journal of Intelligent & Fuzzy Systems* 33 (4). IOS Press: 2523–2531. doi:10.3233/JIFS-17740.
- Yesilnacar, E., and T. Topal. 2005. “Landslide Susceptibility Mapping: A Comparison of Logistic Regression and Neural Networks Methods in a Medium Scale Study, Hendek Region

(Turkey).” *Engineering Geology* 79 (3–4). Elsevier: 251–266.
doi:10.1016/j.enggeo.2005.02.002.

Zhang, Kaixiang, Xueling Wu, Ruiqing Niu, Ke Yang, and Lingran Zhao. 2017. “The Assessment of Landslide Susceptibility Mapping Using Random Forest and Decision Tree Methods in the Three Gorges Reservoir Area, China.” *Environmental Earth Sciences* 76 (11). Springer Berlin Heidelberg: 405. doi:10.1007/s12665-017-6731-5.